MULTIMEDIA UNIVERSITY®

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, **2015/2016**

HPB2031 – Bioinformatics Algorithms 1

(All sections / Groups)

29 FEB 2016
9:00 -11:00 AM
[2 hours]

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 4 pages.
2. You are required to attempt four (4) questions out of five (5) questions. All questions carry equal marks.
3. Please write all your answers in the Answer Booklet provided.

Question 1 [10 marks]

a) Using the dynamic programming method, construct the global alignment score table for the following two sequences. The scoring parameters are:
Match score = +1, mismatch score = -1, gap penalty = -2 [4 marks]
TCGCA
TCCA

b) Find a possible alignment using trace back step. [3 marks]

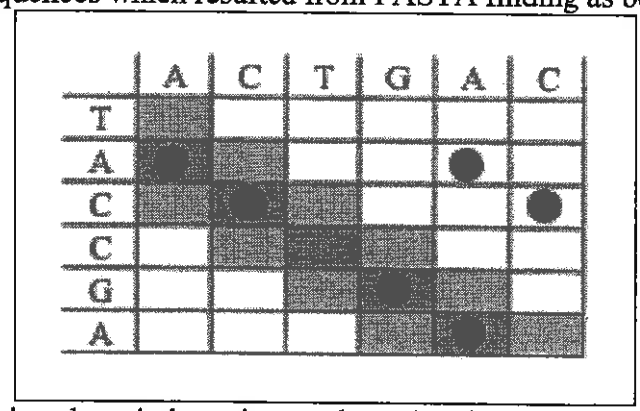
c) Explain the terms "Similarity", "Homology", "Ortholog" and "Paralog".[3 marks]

Question 2 [10 marks]

a) What is the principle behind the scoring for amino acid substitution in sequence alignment? Given one example of substitution matrix. [3 marks]

b) What are the similarities of the logic of BLAST and FASTA methods in sequence analysis? [2 marks]

c) Given both sequences which resulted from FASTA finding as below:



i. Determine the window size or the value for k-tuples for this kind of alignment. [1 mark]

ii. What is the optimal tuple(s) value for DNA and amino acids in FASTA algorithm? [2 marks]

iii. What is the purpose of creating the matrix given above as an initial step in FASTA approach? [2 marks]

Question 3 [10 marks]

a) You have a favourite gene, and you want to determine in what tissue it is expressed. Name one online database you are likely to use to obtain the information. Explain your answer briefly. [2 marks]

Continued...

b) NCBI integrates sequence data from many different sources that name sequence records in different ways. To facilitate database management, different sequence identifiers have been used. Give two sequence identifiers and explain their purposes. [2 marks]

c) You have a protein sequence, and you want to quickly predict its structure. After performing BLAST and PSI-BLAST searches, you identify the most closely related proteins with known structures as several having 15% amino acid identities to your protein, with a non-significant expected value. What will you do? Why? [2 marks]

d) Would one expect secondary structure prediction methods which are intended for soluble proteins to identify membrane helices as helices? Explain. [2 marks]

e) The protein function can be predicted according to the protein sequence and/or structural similarity. Why it is better to use the structural similarity information to study protein function? [2 marks]

Question 4 [10 marks]

a) Describe the Neighbor-Joining method for deriving phylogenetic trees. How does it differ from Distance methods? [3 marks]

b) Given the distance matrix below (N = 6), show the Neighbor-Joining steps to derive the corresponding phylogenetic tree. (Calculate r(i), then new matrix M(i,j), choose neighbors as new 'internal node', calculate distance matrix with new node, …) [7 marks]

	A	B	C	D	E
B	5				
C	4	7			
D	7	10	7		
E	6	9	6	5	
F	8	11	8	9	8

Continued…

Question 5 [10 marks]

a) In the context of Pathways, what is a PGDB? What are the components (tiers) that exist in PGDBs? [4 marks]

b) Give two main differences between "pathway tools" and KEGG. [2 marks]

c) List all kinds of biological pathways or networks and how they are related to "systems biology". [4 marks]

End of Paper